FILE 'BIOSIS, MEDLINE, EMBASE, EMBAL, SCISEARCH, BIOTECHDS, CAPLUS'
ENTERED AT 18:25:33 ON 28 SEP 2001

L1 141 S (**THREE**()DIMENSION?()**DATABASE**?)
L2 192 S L1 OR (3()D()**DATABASE**?)
L3 42 S L2 AND (**PROTEIN**? OR **PEPTIDE**?) AND **STRUCTURE**?
L4 10 S L3 AND (**PREDICT**? OR **DETERMIN**? OR **GUESS**?)
L5 6 DUP REM L4 (4 DUPLICATES REMOVED)
L6 1 S L5 AND (**SEGMENT**?)
L7 5 S L5 NOT L6
L8 402093 S (**PROTEIN**? AND **CHARACTERIZATION**)
L9 1843 S L8 AND (**MODELING**?)
L10 50 S L9 AND (**DATABASE**?)
L11 9 S L10 AND ((3()D) OR (3()**DIMENSIONAL**) OR (**THREE**()**DIMENSIONAL**))
L12 8 DUP REM L11 (1 DUPLICATE REMOVED)
L13 2047 S (**PROTEIN**()**STRUCTURE**?()**PREDICT**?) OR (**PREDICT**?()**PROTEIN**?()**STRUC**
L14 411 S L13 AND (**DATABASE**?)
L15 93 S L14 AND ((3()D) OR (3()**DIMENSIONAL**) OR (**THREE**()**DIMENSIONAL**))
L16 51 DUP REM L15 (42 DUPLICATES REMOVED)
L17 50 S L16 NOT L11

L7 ANSWER 1 OF 5 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1998:490986 BIOSIS
DOCUMENT NUMBER: PREV199800490986
TITLE: **Predict**ion of binding constants of **protein** ligands: A fast method for the prioritization of hits obtained from de novo design or 3D **database** search programs.
AUTHOR(S): Boehm, Hans-Joachim (1)
CORPORATE SOURCE: (1) Hoffmann-La Roche Ltd., Pharmaceuticals Division, Computational Chemistry, CH-4070 Basel Switzerland
SOURCE: Journal of Computer-Aided Molecular Design, (July, 1998) Vol. 12, No. 4, pp. 309-323. ISSN: 0920-654X.
DOCUMENT TYPE: Article
LANGUAGE: English
AB A dataset of 82 **protein**-ligand complexes of known 3D **structure** and binding constant Ki was analysed to elucidate the important factors that **determine** the strength of **protein** -ligand interactions. The following parameters were investigated: the number and geometry of hydrogen bonds and ionic interactions between the **protein** and the ligand, the size of the lipophilic contact surface, the flexibility of the ligand, the electrostatic potential in the

binding site, water molecules in the binding site, cavities along the **protein**-ligand interface and specific interactions between aromatic rings. Based on these parameters, a new empirical scoring function is presented that estimates the free energy of binding for a **protein**-ligand complex of known 3D **structure**. The function distinguishes between buried and solvent accessible hydrogen bonds. It tolerates deviations in the hydrogen bond geometry of up to 0.25 ANG in the length and up to 30degree in the hydrogen bond angle without penalizing the score. The new energy function reproduces the binding constants (ranging from 3.7 X 10-2 M to 1 X 10-14 M, corresponding to binding energies between -8 and -80 kJ/mol) of the dataset with a standard deviation of 7.3 kJ/mol corresponding to 1.3 orders of magnitude in binding affinity. The function can be evaluated very fast and is therefore also suitable for the application in a 3D **database** search or de novo ligand design program such as LUDI. The physical significance of the individual contributions is discussed.

L7 ANSWER 2 OF 5 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1998:50503 BIOSIS
DOCUMENT NUMBER: PREV199800050503
TITLE: Identification of novel farnesyl **protein** transferase inhibitors using **three-dimensional database** searching methods.
AUTHOR(S): Kaminski, James J. (1); Rane, D. F.; Snow, Mark E.; Weber, Lois; Rothofsky, Marnie L.; Anderson, Samantha D.; Lin, Stanley L.
CORPORATE SOURCE: (1) Schering-Plough Res. Inst., Kenilworth, NJ 07033 USA
SOURCE: Journal of Medicinal Chemistry, (Dec. 2, 1997) Vol. 40, No. 25, pp. 4103-4112.
ISSN: 0022-2623.
DOCUMENT TYPE: Article
LANGUAGE: English

AB Generation of a **three-dimensional** pharmacophore model (hypothesis) that correlates the biological activity of a series of farnesyl **protein** transferase (FPT) inhibitors, exemplified by the prototype 1-(4-pyridylacetyl)- 4-(8-chloro-5,6-dihydro-11H-benzo(5,6)cyclohepta(1,2-b)pyridin-11-ylidene)piperidine, Sch 44342, 1, with their chemical **structure** was accomplished using the **three-dimensional** quantitative **structure**-activity relationship (3D-QSAR) software program, Catalyst. On the basis of the in vitro FPT inhibitory activity of a training set of compounds, a five-feature hypothesis containing four hydrophobic and one hydrogen bond acceptor region was generated. Using this hypothesis as a **three-dimensional** query to search our corporate **database** identified 718 compounds (hits). **Determin**ation of the in vitro FPT inhibitory activity using available compounds from this "hitlist" identified five compounds, representing **three struc**turally novel classes, that exhibited in vitro FPT inhibitory activity, IC50 ltoreq 5

muM. From these **three** classes, a series of substituted dihydrobenzothiophenes was selected for further **structure**-FPT inhibitory activity relationship studies. The results from these studies is discussed.

L7 ANSWER 3 OF 5 MEDLINE
ACCESSION NUMBER: 1998020698 MEDLINE
DOCUMENT NUMBER: 98020698 PubMed ID: 9377091
TITLE: The discovery, **characterization** and crystallographically **determin**ed binding mode of an FMOC-containing inhibitor of HIV-1 protease.
AUTHOR: Rutenber E E; De Voss J J; Hoffman L; Stroud R M; Lee K H; Alvarez J; McPhee F; Craik C; Ortiz de Montellano P R
CORPORATE SOURCE: Department of Biochemistry and Biophysics, University of California at San Francisco, 94143, U.S.A.
CONTRACT NUMBER: GM39522 (NIGMS)
SOURCE: BIOORGANIC AND MEDICINAL CHEMISTRY, (1997 Jul) 5 (7) 1311-20.
Journal code: B38; 9413298. ISSN: 0968-0896.
PUB. COUNTRY: ENGLAND: United Kingdom
Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE **SEGMENT**: Priority Journals
ENTRY MONTH: 199711
ENTRY DATE: Entered STN: 19971224
Last Updated on STN: 19971224
Entered Medline: 19971110

AB A pharmacophore derived from the **structure** of the dithiolane derivative of haloperidol bound in the active site of the HIV-1 protease (HIV-1 PR) has been used to search a **three-dimensional database** for new inhibitory frameworks. This search identified an FMOC-protected N-tosyl arginine as a lead candidate. A derivative in which the arginine carboxyl has been converted to an amide has been crystallized with HIV-1 PR and the **structure** has been **determin**ed to a resolution of 2.5 A with a final R-factor of 18.5%. The inhibitor binds in an extended conformation that results in occupancy of the S2, S1', and S3' subsites of the active site. Initial **structure**-activity studies indicate that: (1) the FMOC fluorenyl moiety interacts closely with active site residues and is important for binding; (2) the N(G)-tosyl group is necessary to suppress protonation of the arginine guanidinyl terminus; and (3) the arginine carboxamide function is involved in interactions with the water coordinated to the catalytic aspartyl groups. FMOC-protected arginine derivatives, which appear to be relatively specific and nontoxic, offer promise for the development of useful HIV-1 protease inhibitors.

L7 ANSWER 4 OF 5 SCISEARCH COPYRIGHT 2001 ISI (R)

ACCESSION NUMBER: 1999:316486 SCISEARCH
THE GENUINE ARTICLE: 187WV
TITLE: The discovery of steroids and other novel FKBP inhibitors using a molecular docking program
AUTHOR: Burkhard P; Hommel U; Sanner M; Walkinshaw M D (Reprint)
CORPORATE SOURCE: UNIV EDINBURGH, **STRUCT** BIOCHEM UNIT, MICHAEL SWANN BLDG, KINGS BLDG, EDINBURGH EH9 3JR, MIDLOTHIAN, SCOTLAND (Reprint); UNIV EDINBURGH, **STRUCT** BIOCHEM UNIT, EDINBURGH EH9 3JR, MIDLOTHIAN, SCOTLAND
COUNTRY OF AUTHOR: SCOTLAND
SOURCE: JOURNAL OF MOLECULAR BIOLOGY, (16 APR 1999) Vol. 287, No. 5, pp. 853-858.
Publisher: ACADEMIC PRESS LTD, 24-28 OVAL RD, LONDON NW1 7DX, ENGLAND.
ISSN: 0022-2836.
DOCUMENT TYPE: Article; Journal
FILE **SEGMENT**: LIFE
LANGUAGE: English
REFERENCE COUNT: 24
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The molecular docking computer program SANDOCK was used to screen small molecule **three-dimensional database**s in the hunt for novel FKBP inhibitors. Spectroscopic measurements confirmed binding of over 20 compounds to the target **protein**, some with dissociation constants in the low micromolar range. The discovery that FK506 binding **protein** is a steroid binding **protein** may be of wider biological significance. Two-**dimensional** NMR was used to **determin**e the steroid binding mode and confirmed the interactions **predicte**d by the docking program. (C) 1999 Academic Press.

L7 ANSWER 5 OF 5 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 96:641597 SCISEARCH
THE GENUINE ARTICLE: VD605
TITLE: THE **PROTEIN** DATA-BANK - CURRENT STATUS AND FUTURE CHALLENGES
AUTHOR: ABOLA E E (Reprint); MANNING N O; PRILUSKY J; STAMPF D R; SUSSMAN J L
CORPORATE SOURCE: BROOKHAVEN NATL LAB, DEPT CHEM, UPTON, NY, 11973 (Reprint); WEIZMANN INST SCI, BIOINFORMAT UNIT, IL-76100 REHOVOT, ISRAEL; WEIZMANN INST SCI, DEPT BIOL **STRUCT**,

IL-76100 REHOVOT, ISRAEL; BROOKHAVEN NATL LAB, DEPT BIOL,
UPTON, NY, 11973
COUNTRY OF AUTHOR: USA; ISRAEL
SOURCE: JOURNAL OF RESEARCH OF THE NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY, (MAY/JUN 1996) Vol. 101, No. 3, pp. 231-241.
ISSN: 1044-677X.
DOCUMENT TYPE: Article; Journal
FILE **SEGMENT**: PHYS; ENGI
LANGUAGE: ENGLISH
REFERENCE COUNT: 24
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB **Protein** Data Bank (PDB) is an archive of experimentally **determined three-dimensional structures** of **proteins**, nucleic acids, and other biological macromolecules with a 25 year history of service to a global community. PDB is being replaced by 3DB, the **Three-Dimensional Database** of Biomolecular **Structures** that will continue to operate from Brookhaven National Laboratory. 3DB will be a highly sophisticated knowledge-based system for archiving and accessing **struct**ural information that combines the advantages of object oriented and relational **database** systems. 3DB will operate as a direct-deposition archive that will also accept third-party supplied annotations. Conversion of PDB to 3DB will be evolutionary, providing a high degree of compatibility with existing software.

L17 ANSWER 1 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 2001:307268 BIOSIS
DOCUMENT NUMBER: PREV200100307268
TITLE: Thermodynamic propensities of amino acids in the native state ensemble: Implications for fold recognition.
AUTHOR(S): Wrabl, James O.; Larson, Scott A.; Hilser, Vincent J. (1)
CORPORATE SOURCE: (1) Department of Human Biological Chemistry and Genetics and Sealy Center for **Struc**tural Biology, University of Texas Medical Branch, 5.162 Medical Research Bldg., Galveston, TX, 77555-1055: vince@hbcg.utmb.edu USA
SOURCE: **Protein** Science, (May, 2001) Vol. 10, No. 5, pp. 1032-1045. print.
ISSN: 0961-8368.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English
AB An amino acid sequence, in the context of the solvent environment,

contains all of the thermodynamic information necessary to encode a **three-dimensional protein structure**. To investigate the relationship between an amino acid sequence and its corresponding **protein** fold, a **database** of thermodynamic stability information was assembled that spanned 2951 residues from 44 nonhomologous **protein**s. This information was obtained using the COREX algorithm, which computes an ensemble-based description of the native state of a **protein**. It was observed that amino acid types partitioned unequally into high, medium, and low thermodynamic stability environments. Furthermore, these distributions were reproducible and were significantly different than those expected from random partitioning. To assess the **struc**tural importance of the distributions, simple fold-recognition experiments were performed based on a 3D-1D scoring matrix containing only COREX residue stability information. This procedure was able to recover amino acid sequences corresponding to correct target **structure**s more effectively than scoring matrices derived from randomized data. High-scoring sequences were often aligned correctly with their corresponding target profiles, suggesting that calculated thermodynamic stability profiles have the potential to encode sequence information. As a control, identical fold-recognition experiments were performed on the same **database** of **protein**s using DSSP secondary **structure** information in the scoring matrix, instead of COREX residue stability information. The comparable performance of both approaches suggested that COREX residue stability information and secondary **structure** information could be of equivalent utility in more sophisticated fold-recognition techniques. The results of this work are a consequence of the idea that amino acid sequences fold not into single, rigidly stable **structure**s but rather into thermodynamic ensembles best represented by a time-averaged **structure**.

L17 ANSWER 3 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1999:496274 BIOSIS
DOCUMENT NUMBER: PREV199900496274
TITLE: Analysis and assessment of ab initio **three-dimensional predict**ion, secondary **structure**, and contacts **predict**ion.
AUTHOR(S): Orengo, C. A. (1); Bray, J. E.; Hubbard, T.; LoConte, L.; Sillitoe, I.
CORPORATE SOURCE: (1) Department of Biochemistry and Molecular Biology, University College London, Gower Street, London, WC1E 6BT UK
SOURCE: **Protein**s, (1999) Vol. 0, No. SUPPL. 3, pp. 149-170. ISSN: 0887-3585.
DOCUMENT TYPE: Article
LANGUAGE: English
SUMMARY LANGUAGE: English
AB CASP3 saw a substantial increase in the volume of ab initio 3D **predict**ion

data, with 507 datasets for fifteen selected targets and sixty-one groups participating. As with CASP2, methods ranged from computationally intensive strategies that attempt to recreate the physical and chemical forces involved in **protein** folding to the more recent knowledge-based approaches. These exploit information from the **structure databases**, extracting potentially similar fragments and/or distance constraints derived from multiple sequence alignments. The knowledge-based approaches generally gave more consistently successful **predict**ions across the range of targets, particularly that of the Baker group (Bystroff and Baker, J Mol Biol 1998;281:565-577; Simons et al. **Protein**s Suppl 1999;3:171-176), which used a fragment library. In the secondary **structure predict**ion category, the most successful approaches built on the concepts used in PHD (Rost et al. Comput Appl Biosci 1994;10:53-60), an accepted standard inthis field. Like PHD, they exploit neural networks but have different strategies for incorporating multiple sequence data or position-dependent weight matrices for training the networks. Analysis of the contact data, for which only six groups participated, suggested that as yet this data provides a rather weak signal. However, in combination with other types of **predict**ion data it can sometimes be a useful constraint for identifying the correct **structure**.

L17 ANSWER 4 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1998:400280 BIOSIS
DOCUMENT NUMBER: PREV199800400280
TITLE: **Predict**ion of local **structure** in **protein**s using a library of sequence-**structure** motifs.
AUTHOR(S): Bystroff, Christopher; Baker, David
CORPORATE SOURCE: Dep. Biochem., Univ. Washington, Seattle, WA 98195-7350 USA
SOURCE: Journal of Molecular Biology, (Aug. 21, 1998) Vol. 281, No. 3, pp. 565-577.
ISSN: 0022-2836.
DOCUMENT TYPE: Article
LANGUAGE: English
AB We describe a new method for local **protein structure predict**ion based on a library of short sequence pattern that correlate strongly with **protein three-dimensional struct**ural elements. The library was generated using an automated method for finding correlations between **protein** sequence and local **structure**, and contains most previously described local sequence-**structure** correlations as well as new relationships, including a diverging type-II beta-turn, a frayed helix, and a proline-terminated helix. The query sequence is scanned for **segment**s 7 to 19 residues in length that strongly match one of the 82 patterns in the library. Matching **segments** are assigned the **three-dimensional structure** characteristic of the corresponding sequence pattern, and backbone torsion angles for the entire

query sequence are then **predict**ed by piecing together mutually compatible **segment predict**ions. In **predict**ions of local **structure** in a test set of 55 **protein**s, about 50% of all residues, and 76% of residues covered by high-confidence **predict**ions, were found in eight-residue **segment**s within 1.4 ANG of their-true **structure**s. The **predict**ions are complementary to traditional secondary **structure predict**ions because they are considerably more specific in turn regions, and may contribute to ab initio tertiary **structure predict**ion and fold recognition.

L17 ANSWER 5 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1998:130372 BIOSIS
DOCUMENT NUMBER: PREV199800130372
TITLE: An integrated sequence-**structure database** incorporating matching mRNA sequence, amino acid sequence and **protein three-dimensional structure** data.
AUTHOR(S): Adzhubei, Ivan A.; Adzhubei, Alexei A. (1); Neidle, Stephen
CORPORATE SOURCE: (1) CRC Biomolecular **Structure** Unit, Inst. Cancer Res., Sutton, Surrey SM2 5NG UK
SOURCE: Nucleic Acids Research, (Jan. 1, 1998) Vol. 26, No. 1, pp. 327-331.
ISSN: 0305-1048.
DOCUMENT TYPE: Article
LANGUAGE: English
AB We have con**struct**ed a non-homologous **database**, termed the Integrated Sequence-**Structure Database** (ISSD) which comprises the coding sequences of genes, amino acid sequences of the corresponding **protein**s, their secondary **structure** and variant phi,psi angles assignments, and poly**peptide** backbone coordinates. Each **protein** entry in the **database** holds the alignment of nucleotide sequence, amino acid sequence and the PDB **three-dimensional structure** data. The nucleotide and amino acid sequences for each entry are selected on the basis of exact matches of the source organism and cell environment. The current version 1.0 of ISSD is available on the WWW at http://www.**protein**.bio.msu.su/issd/ and includes 107 non-homologous mammalian **protein**s, of which 80 are human **protein**s. The **database** has been used by us for the analysis of synonymous codon usage patterns in mRNA sequences showing their correlation with the **three-dimensional structure** features in the encoded **protein**s. Possible ISSD applications include optimization of **protein** expression, improvement of the **protein structure predict**ion accuracy, and analysis of evolutionary aspects of the nucleotide sequence-**protein structure** relationship.

L17 ANSWER 6 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1997:438129 BIOSIS

DOCUMENT NUMBER: PREV199799737332
TITLE: Sisyphus and **predict**ion of **protein structure**.
AUTHOR(S): Rost, Burkhard; O'Donoghue, Sean
CORPORATE SOURCE: European Molecular Biol. Lab., **Protein** Design Group, Postfach, Meyerhofstrasse 1, D-69012 Heidelberg Germany
SOURCE: CABIOS, (1997) Vol. 13, No. 4, pp. 345-356.
DOCUMENT TYPE: Article
LANGUAGE: English
AB The problem of **predict**ing **protein structure** from the sequence remains fundamentally unsolved despite more than **three** decades of intensive research effort. However, new and promising methods in **three-dimensional** (3D), 2D and 1D **predict**ion have reopened the field. Mean-force-potentials derived from the **protein databases** can distinguish between correct and incorrect models (3D). Inter-residue contacts (2D) can be detected by analysis of correlated mutations, albeit with low accuracy. Secondary **structure**, solvent accessibility and transmembrane helices (1D) can be **predict**ed with significantly improved accuracy using multiple sequence alignments. Some of these new **predict**ion methods have proven accurate and reliable enough to be useful in genome analysis, and in experimental **structure** **determin**ation. Moreover, the new generation of theoretical methods is increasingly influencing experiments in molecular biology.

L17 ANSWER 8 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1995:106286 BIOSIS
DOCUMENT NUMBER: PREV199598120586
TITLE: **Protein struc**tural similarities **predict**ed by a sequence-**structure** compatibility method.
AUTHOR(S): Matsuo, Yo (1); Nishikawa, Ken
CORPORATE SOURCE: (1) **Protein** Eng. Res. Inst., 6-2-3 Furuedai, Suita, Osaka 565 Japan
SOURCE: **Protein** Science, (1994) Vol. 3, No. 11, pp. 2055-2063. ISSN: 0961-8368.
DOCUMENT TYPE: Article
LANGUAGE: English
AB A method for **protein structure predict**ion has been developed, which evaluates the compatibility of an amino acid sequence with known 3-**dimensional structure**s and identifies the most likely **structure**. The method was applied to a large number of sequences in a **database**, and the **structure**s of the following **protein**s were **predict**ed: (1) shikimate kinase (SKase), (2) the hydrophilic subunit of mannose permease (IIAB-Man), (3) rat tyrosine aminotransferase (Tyr AT), and (4) threonine dehydratase (TDH). The functional and evolutionary implications of the **predict**ions are discussed. (1) The **struc**tural similarity between SKase and adenylate kinase was **predict**ed. Alignment of their sequences reveals that the ATP-binding type

A sequence motif and 2 ATP-binding arginine residues are conserved. The **predict**ion suggests a similarity in their functional mechanisms as well as an evolutionary relationship. (2) The **struc**tural similarity between IIAB-Man and galactose/glucose-binding **protein** (GGBP) was **predicte**d. The IIA and IIB domains are aligned with the N- and C-terminal domains of GGBP, respectively. The 2 phosphorylated residues, His 10 and His 175, of IIABx-Man are threaded onto loops located in the substrate-binding cleft of GGBP. The **predict**ion accounts for the phosphoryl transfer from His 10 to His 175, and to the sugar substrate. (3) The **struc**tural similarity between rat Tyr AT and Escherichia coli aspartate AT was **predicte**d, as well as (4) the **struc**tural similarity between TDH and the tryptophan synthase beta subunit. **Predict**ions (3) and (4) support the previous **predict**ions based on observations of the functional similarities between the **protein**s.

L17 ANSWER 10 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1993:495623 BIOSIS
DOCUMENT NUMBER: PREV199396119630
TITLE: **Predict**ion of **protein structure** by evaluation of sequence-**structure** fitness: Aligning sequences to contact profiles derived from **three-dimensional structure**s.
AUTHOR(S): Ouzounis, Christos; Sander, Chris; Scharf, Michael; Schneider, Reinhard
CORPORATE SOURCE: **Protein** Design Group, EMBL, D-6900 Heidelberg Germany
SOURCE: Journal of Molecular Biology, (1993) Vol. 232, No. 3, pp. 805-825.
ISSN: 0022-2836.
DOCUMENT TYPE: Article
LANGUAGE: English
AB The problem of **protein structure predict**ion is formulated here as that of evaluating how well an amino acid sequence fits a hypothetical **structure**. The simplest and most complicated approaches, secondary **structure predict**ion and all-atom free energy calculations, can be viewed as sequence-**structure** fitness problems. Here, an approach of intermediate complexity is described, which involves; (1) description of a **protein structure** in terms of contact interface vectors, with both intra-**protein** and **protein**-solvent contacts (3) generation of numerous hypothetical model **structures** by placing the input sequence into a large set of known **three-dimensional structures** in all possible alignments, (4) evaluation of these models by summing the sequence preferences over all **struc**tural dependent core weights derived from multiple sequence alignments. A number of tests of the method are performed: (1) evaluation of cyclic shifts of a sequence in its native **structure**; (2) alignment of a sequence in its native **structure**, allowing gaps; (3) alignment search with a sequence or sequence fragment in a

database of **structures**; and (4) alignment search with a **structure** in a **database** of sequences. The main results are: (1) a native sequence can very well find its native **structure** among a large number of alternatives, in correct alignment; (2) sub**structure**s, such as (beta-alpha)-n units, can be detected in spite of very low sequence similarity; (3) remote homologues can be detected, with some dependence on the set of parameters used; (4) contact interface parameters are clearly superior to classical secondary **structure** parameters; (5) a simple interface description in terms of just two states, **protein-protein** and **protein**-water contacts, performs surprisingly well; (6) the use of core weights considerably improves accuracy in detection of remote homologues; (7) based on a sequence **database** search with a myoglobin contact profile, the C-terminal domain of a viral origin of replication binding **protein** is **predicte**d to have an all-helical fold. The sequence-**structure** fitness concept is sufficiently general to accommodate a large variety of **protein structure predict**ion methods, including new models of intermediate complexity currently being developed.

L17 ANSWER 11 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS
ACCESSION NUMBER: 1993:384026 BIOSIS
DOCUMENT NUMBER: PREV199396059326
TITLE: Learning and alignment methods applied to **protein structure predict**ion.
AUTHOR(S): Gracy, J. (1); Chiche, L.; Sallantin, J. (1)
CORPORATE SOURCE: (1) Laboratoire Informatique, Robotique Micro-Electronique Montpellier, 860 rue de St. Priest, 34090 Montpellier France
SOURCE: Biochimie (Paris), (1993) Vol. 75, No. 5, pp. 353-361. ISSN: 0300-9084.
DOCUMENT TYPE: Article
LANGUAGE: English
AB Learning techniques are able to extract **struc**tural knowledge specific to a selected set of **protein**s. We describe two algorithms that optimize scores expressing the propensity of a poly**peptide** sequence to adopt a local fold. The first algorithm generates secondary **structure predict**ion rules based on a dictionary of geometrical patterns frequently found in the learning **database**. The second algorithm leads to scores that indicate the fit between an amino acid and a given local **struc**tural environment. Dynamic programming is then used to align **struc**tural information profiles by modifying the local mutation cost with the above learned functions. The main features of the system are exemplified on the **struc**tural **predict**ion of the N-terminal domain of the CD4 antigen. Then the usefulness of additional 3-D information in the alignment is benchmarked on eight pairs of weakly homologous **protein**s.

L17 ANSWER 12 OF 50 BIOSIS COPYRIGHT 2001 BIOSIS

ACCESSION NUMBER: 1991:132440 BIOSIS
DOCUMENT NUMBER: BA91:68980
TITLE: **DATABASE** OF HOMOLOGY-DERIVED **PROTEIN STRUCTURES** AND THE **STRUC**TURAL MEANING OF SEQUENCE ALIGNMENT.
AUTHOR(S): SANDER C; SCHNEIDER R
CORPORATE SOURCE: EUROPEAN MOLECULAR BIOL. LAB., POSTFACH 10-2209, MEYERHOFSTR. 1, D-6900 HEIDELBERG, W. GER.
SOURCE: **PROTEINS STRUC**T FUNCT GENET, (1991) 9 (1), 56-68.
CODEN: PSFGEY. ISSN: 0887-3585.
FILE **SEGMENT**: BA; OLD
LANGUAGE: English
AB The **database** of known **protein three-dimensional structure**s can be significantly increased by the use of sequence homology, based on the following observations. (1) The **database** of known sequences, currently at more than 12,000 **protein**s, is two orders of magnitude larger than the **database** of known **structure**s. (2) The currently most powerful method of **predict**ing **protein structure**s is model building by homology. (3) **Struc**tural homology can be inferred from the level of sequence similarity. (4) The threshold of sequence similarity sufficient for **struc**tural homology depends strongly on the length of the alignment. Here, we first quantify the relation between sequence similarity, **structure** similarity, and alignment length by an exhaustive survey of alignments between **protein**s of known **structure** and report a homology threshold curve as a function of alignment length. We then produce a **database** of homology-derived secondary **structure** of **protein**s (HSSP) by aligning to each **protein** of known **structure** all sequences deemed homologous on the basis of the threshold curve. For each known **protein structure**, the derived **database** contains the aligned sequences, secondary **structure**, sequence variability, and sequence profile. Tertiary **structure**s of the aligned sequences are implied, but not modeled explicitly. The **database** effectively increases the number of known **protein structure**s by a factor of five to more than 1800. The results may be useful in assessing the **struc**tural significance of matches in sequence **database** searches, in deriving preferences and patterns for **structure predict**ion, in elucidating the **struc**tural role of conserved residues, and in **modeling three-dimensional** detail by homology.

L17 ANSWER 17 OF 50 EMBASE COPYRIGHT 2001 ELSEVIER SCI. B.V.
ACCESSION NUMBER: 93207592 EMBASE
DOCUMENT NUMBER: 1993207592
TITLE: Inverted **protein structure predict**ion.

AUTHOR: Bowie J.U.; Eisenberg D.
CORPORATE SOURCE: Molecular Biology Institute, Univ of California at Los Angeles, 405 Hilgard Avenue,Los Angeles, CA 90024-1517, United States
SOURCE: Current Opinion in **Struc**tural Biology, (1993) 3/3 (437-444).
ISSN: 0959-440X CODEN: COSBEF
COUNTRY: United Kingdom
DOCUMENT TYPE: Journal; General Review
FILE **SEGMENT**: 027 Biophysics, Bioengineering and Medical Instrumentation
029 Clinical Biochemistry
LANGUAGE: English
SUMMARY LANGUAGE: English
AB Today we know of over 1000 **protein structure**s, which can be classified into approximately 120 distinct folding patterns. The **database** of known **structure**s provides numerous examples of **protein**s that adopt very similar folds, with some in each folding class having similar sequences. But there are also examples of **protein**s with similar **structure**s that share no obvious sequence similarity. Thus among the 60 000 known amino acid sequences, there must be many that adopt the 120 known folds but cannot be identified based on sequence relationships alone. It is the goal of inverted **protein structure predict**ion to **determine** whether an amino acid sequence adopts a known **structure**. Here, we review the recent, rapid progress in inverted **structure predict**ion. The power of this new generation of methods is that, instead of looking for similarity in sequences, they attempt to match one-**dimensional** sequences directly to **three-dimensional** folds.

L17 ANSWER 19 OF 50 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 2001:176570 SCISEARCH
THE GENUINE ARTICLE: 403PK
TITLE: **Protein structure predict**ion
AUTHOR: Al-Lazikani B (Reprint); Jung J; Xiang Z X; Honig B
CORPORATE SOURCE: Columbia Univ, Howard Hughes Med Inst, Dept Biochem & Mol Biophys, 630 W 168th St, New York, NY 10032 USA (Reprint); Columbia Univ, Howard Hughes Med Inst, Dept Biochem & Mol Biophys, New York, NY 10032 USA
COUNTRY OF AUTHOR: USA
SOURCE: CURRENT OPINION IN CHEMICAL BIOLOGY, (FEB 2001) Vol. 5, No. 1, pp. 51-56.
Publisher: CURRENT BIOLOGY LTD, 84 THEOBALDS RD, LONDON WC1X 8RR, ENGLAND.
ISSN: 1367-5931.

DOCUMENT TYPE: General Review; Journal
LANGUAGE: English
REFERENCE COUNT: 52
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The **predict**ion of **protein structure**, based primarily on sequence and **structure** homology, has become an increasingly important activity. Homology models have become more accurate and their range of applicability has increased. Progress has come, in part, from the flood of sequence and **structure** information that has appeared over the past few years, and also from improvements in analysis tools. These include profile methods for sequence searches, the use of **three-dimensional structure** information in sequence alignment and new homology **modeling** tools, specifically in the **predict**ion of loop and side-chain conformations. There have also been important advances in understanding the physical chemical basis of **protein** stability and the corresponding use of physical chemical potential functions to identify correctly folded from incorrectly folded **protein** conformations.

L17 ANSWER 21 OF 50 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 2000:596791 SCISEARCH
THE GENUINE ARTICLE: 339XV
TITLE: **Protein** threading using PROSPECT: Design and evaluation
AUTHOR: Xu Y (Reprint); Xu D
CORPORATE SOURCE: OAK RIDGE NATL LAB, DIV LIFE SCI, COMPUTAT BIOSCI SECT,
1060 COMMERCE PK DR, OAK RIDGE, TN 37830 (Reprint)
COUNTRY OF AUTHOR: USA
SOURCE: **PROTEINS-STRUCTURE** FUNCTION AND GENETICS, (15 AUG 2000)
Vol. 40, No. 3, pp. 343-354.
Publisher: WILEY-LISS, DIV JOHN WILEY & SONS INC, 605 THIRD AVE, NEW YORK, NY 10158-0012.
ISSN: 0887-3585.
DOCUMENT TYPE: Article; Journal
FILE **SEGMENT**: LIFE
LANGUAGE: English
REFERENCE COUNT: 27
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The computer system PROSPECT for the **protein** fold recognition using the threading method is described and evaluated in this article. For a given target **protein** sequence and a template **structure**, PROSPECT guarantees to find a globally optimal threading alignment between the two, The scoring function for a threading alignment employed in PROSPECT consists of four additive terms: i) a mutation term, ii) a singleton fitness term, iii) a pairwise-contact potential term, and iv) alignment gap penalties, The current version of PROSPECT considers pair contacts only between core

(alpha-helix or beta-strand) residues and alignment gaps only in loop regions, PROSPECT finds a globally optimal threading efficiently when pairwise contacts are considered only between residues that are spatially close (7 Angstrom or less between the C-beta atoms in the current implementation). On a test set consisting of 137 pairs of target-template **protein**s, each pair being from the same superfamily and having sequence identity less than or equal to 30%, PROSPECT recognizes 69% of the templates correctly and aligns 66% of the **struc**turally alignable residues correctly. These numbers may be compared with the 55% fold recognition and 64% alignment accuracy for the same test set using only scoring terms i), ii), and (iv), indicating the significant contribution from the contact term, The fold recognition and alignment accuracy are further improved to 72% and 74%, respectively, when the secondary **structure** information **predict**ed by the PHD program is used in scoring, PROSPECT also allows a user to incorporate constraints about a target **protein**, e.g., disulfide bonds, active sites, and NOE distance restraints, into the threading process. The system rigorously finds a globally optimal threading under the specified constraints. Test results have shown that the constraints can further improve the performance of PROSPECT, **Protein**s 2000;40:343-354.

L17 ANSWER 22 OF 50 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 2000:469365 SCISEARCH
THE GENUINE ARTICLE: 325GC
TITLE: **Protein structure predict**ion
in the postgenomic era
AUTHOR: Jones D T (Reprint)
CORPORATE SOURCE: BRUNEL UNIV, DEPT BIOL SCI, UXBRIDGE UB8 3PH, MIDDX,
ENGLAND (Reprint)
COUNTRY OF AUTHOR: ENGLAND
SOURCE: CURRENT OPINION IN **STRUC**TURAL BIOLOGY, (JUN 2000) Vol. 10,
No. 3, pp. 371-379.
Publisher: CURRENT BIOLOGY LTD, 84 THEOBALDS RD, LONDON WC1X 8RR, ENGLAND.
ISSN: 0959-440X.
DOCUMENT TYPE: Article; Journal
FILE **SEGMENT**: LIFE
LANGUAGE: English
REFERENCE COUNT: 55
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB As the number of completely sequenced genomes rapidly increases, the postgenomic problem of gene function identification becomes ever more pressing. **Predict**ing the **structure**s of **protein**s encoded by genes of interest is one possible means to glean subtle clues as to the functions

of these **proteins**. There are limitations to this approach to gene identification and a survey of the expected reliability of different **protein structure predict**ion techniques has been undertaken.

L17 ANSWER 25 OF 50 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 1999:954399 SCISEARCH
THE GENUINE ARTICLE: 262VB
TITLE: **Predict**ing **protein three-dimensional structure**
AUTHOR: Moult J (Reprint)
CORPORATE SOURCE: UNIV MARYLAND, MARYLAND BIOTECHNOL INST, CTR ADV RES BIOTECHNOL, 9600 GUDELSKY DR, ROCKVILLE, MD 20850 (Reprint)
COUNTRY OF AUTHOR: USA
SOURCE: CURRENT OPINION IN BIOTECHNOLOGY, (DEC 1999) Vol. 10, No. 6, pp. 583-588.
Publisher: CURRENT BIOLOGY LTD, 34-42 CLEVELAND STREET, LONDON W1P 6LE, ENGLAND.
ISSN: 0958-1669.
DOCUMENT TYPE: General Review; Journal
FILE **SEGMENT**: LIFE
LANGUAGE: English
REFERENCE COUNT: 53
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The current state of the art in **modeling protein structure** has been assessed, based on the results of the GASP (Critical Assessment of **protein Structure Predict**ion) experiments. In comparative **modeling**, improvements have been made in sequence alignment, sidechain orientation and loop building. Refinement of the models remains a serious challenge. Improved sequence profile methods have had a large impact in fold recognition. Although there has been some progress in alignment quality, this factor still limits model usefulness. In ab initio **structure predict**ion, there has been notable progress in building approximately correct **structure**s of 40-60 residue-long **protein** fragments. There is still a long way to go before the general ab initio **predict**ion problem is solved. Overall, the field is maturing into a practical technology, able to deliver useful models for a large number of sequences.

L17 ANSWER 41 OF 50 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 96:32113 SCISEARCH
THE GENUINE ARTICLE: TL573
TITLE: ARE **DATABASE**-DERIVED POTENTIALS VALID FOR SCORING BOTH FORWARD AND INVERTED **PROTEIN**-FOLDING

AUTHOR: ROOMAN M J (Reprint); WODAK S J
CORPORATE SOURCE: FREE UNIV BRUSSELS, UNITE CONFORMAT MACROMOLEC BIOL, AVE
PAUL HEGER, CP 160-16, B-1050 BRUSSELS, BELGIUM (Reprint)
COUNTRY OF AUTHOR: BELGIUM
SOURCE: **PROTEIN** ENGINEERING, (SEP 1995) Vol. 8, No. 9, pp. 849-858

ISSN: 0269-2139.
DOCUMENT TYPE: General Review; Journal
FILE **SEGMENT**: LIFE
LANGUAGE: ENGLISH
REFERENCE COUNT: 82

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB **Database**-derived potentials, compiled from frequencies of sequence and **structure** features, are often used for scoring the compatibility of **protein** sequences and conformations. It is often believed that these scores correspond to differences in free energy with, in addition, a term containing the partition function of the system. Since this function does not depend on the conformation, the potentials are considered to be valid for scoring the compatibility of different conformations with a given sequence ('forward folding'), but not of sequences with a given **structure** ('inverted folding'). This interpretation is questioned here. It is argued that when many body-effects, which dominate frequencies compiled from the **protein database**, are corrected for, the potentials approximate a physically meaningful free energy difference from which the partition function term cancels out. It is the difference between the free energy of a given sequence in a specific conformation and that of the same sequence in a denatured-like state. Two examples of denatured-like states are discussed. Depending on the considered state, the free energy difference reduces to the commonly used scoring scheme, or contains additional terms that depend on the sequence, In both cases, all the terms can be derived from sequence-**structure** frequencies in the **database**. Such free energy difference, commonly defined as the folding free energy, is a measure of **protein** stability and can be used for scoring both forward and inverted **protein** folding. The implications for the use of knowledge-based potentials in **protein structure predict**ion are described. Finally, the difficulty of designing tests that could validate the proposed approach, and the inherent limitations of such tests, are discussed.

L17 ANSWER 44 OF 50 SCISEARCH COPYRIGHT 2001 ISI (R)
ACCESSION NUMBER: 93:544573 SCISEARCH
THE GENUINE ARTICLE: LV383
TITLE: OPTIMAL NEURAL NETWORKS FOR **PROTEIN-STRUCTURE PREDICT**ION

AUTHOR: HEADGORDON T (Reprint); STILLINGER F H
CORPORATE SOURCE: LAWRENCE BERKELEY LAB, BERKELEY, CA, 94720 (Reprint); AT&T
BELL LABS, MURRAY HILL, NJ, 07974
COUNTRY OF AUTHOR: USA
SOURCE: PHYSICAL REVIEW E, (AUG 1993) Vol. 48, No. 2, pp. 1502-1515.
ISSN: 1063-651X.
DOCUMENT TYPE: Article; Journal
FILE **SEGMENT**: PHYS
LANGUAGE: ENGLISH
REFERENCE COUNT: 32
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The successful application of neural-network algorithms for **predict**ion of **protein structure** is stymied by **three** problem areas: the sparsity of the **database** of known **protein structure**s, poorly devised network architectures which make the input-output mapping opaque, and a global optimization problem in the multiple-minima space of the network variables. We present a simplified poly**peptide** model residing in two dimensions with only two amino-acid types, A and B, which allows the **determin**ation of the global energy **structure** for all possible sequences of pentamer, hexamer, and heptamer lengths. This model simplicity allows us to compile a complete **struc**tural **database** and to devise neural networks that reproduce the tertiary **structure** of all sequences with absolute accuracy and with the smallest number of network variables. These optimal networks reveal that the **three** problem areas are convoluted, but that thoughtful network designs can actually deconvolute these detrimental traits to provide network algorithms that genuinely impact on the ability of the network to generalize or learn the desired mappings. Furthermore, the two-**dimensional** poly**peptide** model shows sufficient chemical complexity so that transfer of neural-network technology to more realistic **three-dimensional protein**s is evident.

L17 ANSWER 48 OF 50 CAPLUS COPYRIGHT 2001 ACS
ACCESSION NUMBER: 1997:122928 CAPLUS
DOCUMENT NUMBER: 126:196958
TITLE: **Protein** sequence alignment and **database** scanning
AUTHOR(S): Barton, Geoffrey J.
CORPORATE SOURCE: Laboratory of Molecular Biophysics, University of Oxford, Oxford, OX1 3QU, UK
SOURCE: **Protein Struct. Predict.** (1996), 31-63. Editor(s): Sternberg, Michael J. E. IRL Press: Oxford, UK.
CODEN: 63ZTA7
DOCUMENT TYPE: Conference; General Review
LANGUAGE: English

AB A review and discussion with 69 refs. In the context of **protein structure predict**ion, there are 2 principle reasons for comparing and aligning **protein** sequences: (1) to obtain an accurate alignment which may be for **protein modeling** by comparison to **protein**s of known 3-**dimensional structure** and (2) to scan a **database** with a newly detd. **protein** sequence and identify possible functions for the **protein** by analogy with well-characterized **protein**s. In this chapter, I review the underlying principles and techniques for sequence comparison as applied to **protein**s and used to satisfy these 2 aims.

L17 ANSWER 49 OF 50 CAPLUS COPYRIGHT 2001 ACS
ACCESSION NUMBER: 1995:972162 CAPLUS
DOCUMENT NUMBER: 124:50091
TITLE: **Predict**ion of **protein struct**ural similarities using a 3D-1D compatibility method
AUTHOR(S): Matsuo, Yo; Nishikawa, Ken
CORPORATE SOURCE: **Protein** Engineering Research Institute, Suita, 565, Japan
SOURCE: Genome Inf. Ser. (1994), 5(Genome Informatics Workshop 1994), 11-18
CODEN: GINSE9; ISSN: 0919-9454
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The 3D-1D compatibility method is a new approach to **protein structure predict**ion. It evaluates the compatibility of a one-**dimensional** (1D) amino acid sequence with known **three-dimensional** (3D) **structure**s, and select the most likely **structure**. We have developed a method, which evaluates the 3D-1D compatibility using the following functions: side-chain packing, solvation, hydrogen-bonding, and local conformation functions. The method has been applied to a large no. of sequences in **database**s. Here, the **predict**ions of the **struct**ural similarities between the following pairs are described in detail: spermidine/putrescine-binding **protein** and maltose-binding **protein**, shikimate kinase and adenylate kinase, and mannose permease hydrophilic subunit (IIABMan) and galactose/glucose-binding **protein**. Functional and evolutionary implications of the **predict**ions are discussed. Through these examples of **predict**ions, the present work demonstrates the promise of the 3D-1D method.